

Ten simple rules for reading randomized controlled trials

Diez reglas simples para leer ensayos clínicos aleatorizados

Alvaro Taype-Rondan^{1,2}, Sergio Goicochea-Lugo^{1,2}

¹ Unidad de Investigación para la Generación y Síntesis de Evidencias en Salud, Vicerrectorado de Investigación, Universidad San Ignacio de Loyola, Lima, Peru.
² EviSalud – Evidencias en Salud, Lima, Peru.

Article history

Received: April 21, 2026
Accepted: May 30, 2026
Online: June 8, 2026

How to cite this article

Taype-Rondan A, Goicochea-Lugo S. Ten simple rules for reading randomized controlled trials. *Med Educ Clin Pract.* 2026;1(2):e022. doi: 10.21142/mecp.2026.1.2.e022.

*Correspondence:

Alvaro Taype-Rondan
E-mail: alvaro.taype.r@gmail.com



This work is licensed under a Creative Commons Attribution 4.0 International License

ABSTRACT

Randomized controlled trials (RCTs) are central to evidence-based medicine, yet their interpretation can be challenging for clinicians without formal training in research methods. This article presents ten practical rules to guide the reading, interpretation, and application of RCTs, focusing on common pitfalls and key concepts. These include clarifying the study question using the PICO framework, distinguishing primary from secondary outcomes, understanding the purpose and limits of randomization, interpreting effect sizes and confidence intervals, and considering absolute as well as relative measures. The rules also emphasize systematic assessment of risk of bias, evaluation of external validity, and careful consideration of study limitations and their direction of bias. Finally, readers are encouraged to contextualize findings within the broader body of evidence before applying them to practice. Aimed at students and clinicians beginning to appraise biomedical evidence, these rules are intended as an accessible starting point to support more informed and critical interpretation of RCTs.

Keywords: Randomized Controlled Trials as Topic; GRADE Approach; Evidence-based medicine; Bias; Decision Making (Source: MeSH NLM).

RESUMEN

Los ensayos clínicos aleatorizados (ECA) constituyen un pilar fundamental de la medicina basada en evidencias; sin embargo, su interpretación puede representar un desafío para clínicos sin formación formal en metodología de investigación. Este artículo presenta diez reglas prácticas para orientar la lectura, interpretación y aplicación de los ECA, con énfasis en errores frecuentes y conceptos clave. Entre estos se incluyen la clarificación de la pregunta del ensayo mediante el marco PICO, la diferenciación entre desenlaces primarios y secundarios, la comprensión del propósito y las limitaciones de la aleatorización, la interpretación de las medidas de efecto y de los intervalos de confianza, así como la consideración de medidas de efecto absolutas y relativas. Las reglas también enfatizan la evaluación sistemática del riesgo de sesgo, la valoración de la validez externa y el análisis cuidadoso de las limitaciones del estudio y de la dirección potencial del sesgo. Finalmente, se alienta a los lectores a contextualizar los hallazgos dentro del conjunto global de la evidencia antes de aplicarlos en la práctica. Dirigidas a estudiantes y clínicos que se inician en la evaluación crítica de la evidencia biomédica, estas reglas buscan servir como un punto de partida accesible para promover una interpretación más informada y crítica de los ECA.

Palabras clave: Ensayos Clínicos Aleatorios; Medicina Basada en Evidencias; Uso de la Información Científica en la Toma de Decisiones en Salud (fuente: DeCS Bireme).

INTRODUCTION

Evidence-based medicine rests on the ability to find, evaluate, and apply research evidence to clinical decisions [1]. Among study designs, the randomized controlled trial (RCT) occupies a privileged position: by randomly allocating participants to treatment or control conditions, it minimizes confounding and allows for causal inference [2,3]. It remains the design of choice for regulatory drug approval and for answering intervention questions in clinical practice guidelines.

Despite their importance, RCTs are not always easy to interpret. A typical trial report is filled with methodological and statistical concepts, such as allocation concealment, statistical power, relative risks, confidence intervals, and p-values [3]. For clinicians trained primarily in bedside reasoning rather than research methods, navigating this complexity can be challenging.

The gap between evidence generation and evidence uptake is not merely an academic concern. Misreading a trial (overestimating benefit, missing a key bias, or applying results to patients unlike those studied) has direct consequences for patient care [4,5]. At the same time, rejecting trial evidence simply because its methods appear unclear is equally problematic.

These ten rules are offered as practical guidance. They do not replace formal training in epidemiology or statistics; they are a starting point. Each rule targets a specific, recurring challenge that readers without a robust methodological background encounter when approaching an RCT for the first time.

Our aim is to provide a practical framework for reading/interpreting an individual trial and for judging whether and how its findings might apply to the reader's own clinical question. Thus, these rules move between appraising a trial on its own terms and weighing its relevance to practice. The intended audience is students and clinicians who are beginning to appraise biomedical evidence and who lack extensive formal training in research methods. The scope is intentionally limited to the conventional two-arm, parallel-group superiority trial.

Readers seeking more in-depth guidance may benefit from consulting the Consolidated Standards of Reporting Trials (CONSORT) statement [6], the Cochrane Risk of Bias tool [7], and concise critical appraisal resources such as the Critical Appraisal Skills Programme (CASP) checklist [8].

Ten Simple Rules

Rule 1: Start with the study question, not the conclusions

Before reading the results or conclusions, ask: What question is this trial trying to answer? The PICO framework

(Population, Intervention, Comparator, and Outcome) provides a reliable structure for this task [9]. Identifying each PICO element helps to prevent the most common reading error: anchoring on the headline result without understanding the context in which it was generated.

This is important because the question you have in your clinical practice does not always align with the question the RCT is designed to answer. Ask whether the study population (as defined by its inclusion and exclusion criteria, and by the baseline characteristics of those enrolled) matches your patients, whether the intervention is one you would realistically use (including its dosage, duration, and co-interventions), if the comparator makes sense for your context, and whether the measured outcomes are relevant to your clinical question (Table 1). A trial may be methodologically impeccable and its results entirely irrelevant to your practice.

Rule 2: Distinguish the primary outcome from secondary outcomes and from those that are important to patients

Every well-designed RCT pre-specifies a primary outcome. This is the outcome (or, occasionally, the small set of outcomes) for which the trial is primarily powered and to which the main inferential analyses are directed [10]. However, some trials designate more than one primary outcome, and a few do not clearly specify one. Secondary outcomes are usually not the basis for confirmatory inference, although they may still be pre-specified and clinically meaningful.

If an RCT does not demonstrate an effect on the primary outcome but reports a statistically significant effect on a secondary outcome, the appropriate interpretation is that the trial did not confirm the hypothesized effect. The secondary finding should be considered largely exploratory and hypothesis-generating, rather than confirmatory [11].

On the other hand, the study's primary outcome may not be the most clinically meaningful outcome for patients. For example, in Table 1, a reduction in HbA1c may matter little to a patient who is far more concerned about outcomes such as death, nephropathy, foot ulcers, or cardiovascular events. Thus, even when findings on secondary outcomes are merely hypothesis-generating rather than confirmatory, they can still provide valuable insight into what matters most to patients, helping to guide decision-making.

Finally, a comprehensive interpretation of an RCT requires attention not only to efficacy outcomes but also to safety outcomes (i.e., adverse events and complications associated with the intervention), ensuring a balanced assessment of benefits and risks.

Table 1. The PICO Framework Applied to an RCT

Component	Question	Example in an RCT
P – Population	Who are the participants?	Adults ≥18 years with type 2 diabetes
I – Intervention	What is being tested?	Empagliflozin 10 mg/day
C – Comparator	What is it compared to?	Placebo
O – Outcomes	What is being measured?	Primary outcome: HbA1c reduction at 6 months Secondary outcomes: Death at 6 months; onset of diabetic nephropathy within 6 months; onset of diabetic foot ulcers within 6 months; cardiovascular events (myocardial infarction, stroke) within 6 months; adverse events

Rule 3: Understand what randomization does (and does not) guarantee

Randomization is the feature that most distinguishes the RCT from observational studies. By randomly assigning participants to intervention or control it tends to balance known and unknown confounders evenly across groups, making the two groups comparable at baseline (at the beginning of the study) [2].

A common source of confusion is that randomization is not the same as random sampling. Random sampling refers to how participants are selected from a population to enter a study; randomization refers to how those already enrolled are assigned to groups. Most RCTs recruit volunteers rather than random population samples (and that is acceptable, because the purpose of randomization is not generalizability but internal validity: ensuring that the comparison between groups is fair).

To appreciate what randomization achieves, consider the alternative. In observational studies, patients who receive a treatment are often systematically different from those who do not, making it difficult to separate the effect of the treatment from the effect of those pre-existing differences. Randomization dissolves this problem by making group assignment unpredictable.

However, randomization alone does not guarantee a high-quality trial. A small sample may randomize badly by chance; an unconcealed allocation sequence can be manipulated; and even a well-randomized trial can be compromised by other study limitations [12]. In practice, look beyond the word “randomized”: check how the allocation sequence was generated (for example, by computer-generated random numbers) and, crucially, whether it was concealed from the people enrolling participants, since an unconcealed sequence can be subverted and undo the benefit of randomization [12].

Rule 4: Look at the CONSORT flow diagram

The CONSORT statement requires that authors provide a flow diagram showing the number of participants screened, eligible, randomized, allocated, followed up,

and analyzed [6]. This diagram, usually the first figure in a trial report, contains a remarkable density of information (Figure 1).

Pay attention to the gap between randomized and analyzed participants. Large losses to follow-up threaten the validity of results, particularly when attrition differs between groups in frequency or in the reasons for dropout [7,13]. Although a figure of around 20% is sometimes cited as a rule of thumb, no universal threshold exists: the impact of missing data depends on the type of outcome, the proportion lost, and the mechanism behind the loss [7]. Substantial loss may reduce the effective sample size required to reliably assess the primary outcome and may also compromise the balance achieved through randomization.

The diagram also reveals whether the analysis was conducted as intention-to-treat (all randomized participants regardless of adherence) or per-protocol (only those who completed the intervention), a distinction with substantial implications for how results should be interpreted [13]. Intention-to-treat analyses preserve the benefits of randomization and better reflect real-world effectiveness, whereas per-protocol analyses may better estimate the effect of the intervention under ideal conditions, albeit at the cost of increased susceptibility to bias.

Rule 5: Do not interpret the p-value in isolation

The p-value is probably the most misunderstood number in clinical research. A p-value of 0.04 means, roughly, that if the null hypothesis were true, data as extreme as those observed would arise by chance approximately 4% of the time [14]. Another common but flawed interpretation is that there is a 4% probability that the null hypothesis (i.e., that the intervention does not affect the outcome) is true.

The p-value is highly sensitive to sample size, so a very large trial can yield a statistically significant result for a trivial difference, while a small trial may miss an important effect for lack of power.

The p-value should be read together with two complementary quantities: the effect estimate (expressed

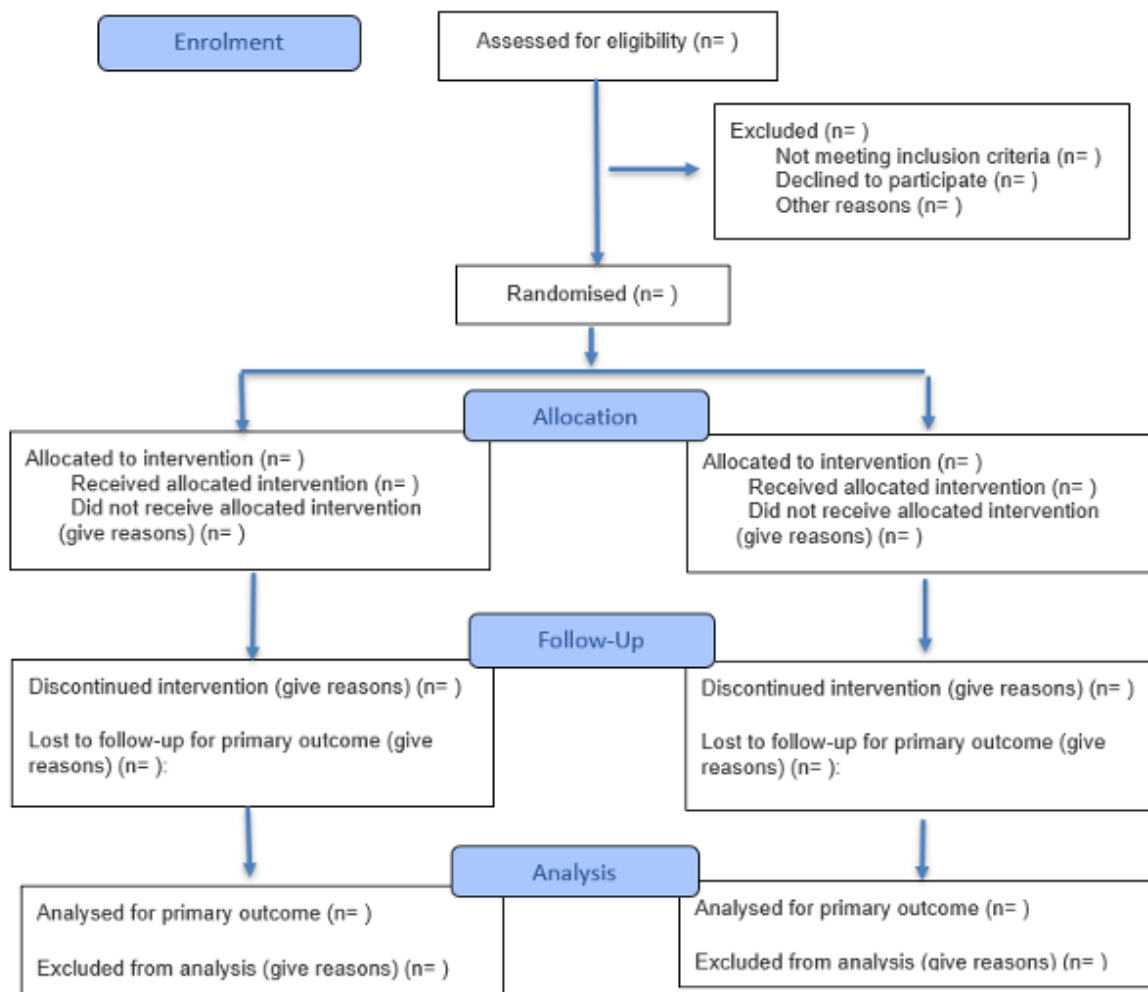


Figure 1. CONSORT flow diagram (template adapted by the authors from the CONSORT 2010 statement^[6])

Table 2. Key concepts and their clinical interpretation

Concept	Plain-language definition	Practical tip
p-value	A measure of how compatible the observed data are with the null hypothesis. It can take values from 0 to 1, and smaller values indicate less compatibility.	$p < 0.05$, usually known as “statistically significant”, does not prove clinical importance; interpret alongside effect size.
Relative risk (RR)	Risk of an event in the treatment group divided by the risk of the same event in the control group.	RR=0.75 means a 25% relative reduction in the event; whether this is favorable depends on the outcome (e.g., desirable for mortality, undesirable for a beneficial response).
95% confidence interval (95% CI)	A range of values that shows where the true effect or a result (e.g., a RR) is likely to be. If the study were repeated many times, about 95 out of 100 confidence intervals would contain the true value.	A CI that crosses zero (for differences) or one (for ratios) indicates a lack of statistical significance; a wide interval signals an imprecise estimate.
Risk difference (RD) or attributable risk (AR)	Difference in event rates between groups.	RD indicates how many individuals will benefit from or be harmed by an intervention. Thus, it is more informative than RR alone.
Number needed to treat (NNT)	Patients needed to be treated to prevent (or cause) one additional outcome.	$NNT = 1 / RD$.

as a relative or absolute measure), which conveys the magnitude and direction of the association, and its confidence interval, which conveys the precision of the estimate (Table 2).

The p-value reflects the outcome of the hypothesis test, whereas the effect estimate and its 95% confidence interval describe how large and how precise that effect is; the two are complementary rather than hierarchical.

For the same reason, results with p-values close to the conventional threshold (for example, 0.049 versus 0.051) should not be interpreted as fundamentally different.

Rule 6: Think in absolute numbers, not only relative ones

Relative measures of effect (such as relative risk, odds ratio, or hazard ratio) are essential for comparing trials but can distort the perceived benefit of a treatment. A 50% relative risk reduction sounds impressive; however, if the absolute event rate falls from 2% to 1% (risk difference [RD] of 1%), the number needed to treat (NNT) is 100, meaning 100 patients must be treated for one additional patient to benefit [15]. Whether this is worthwhile depends on the decision-making perspective: for a clinician or health-system, an NNT of 100 must be weighed against the intervention’s cost, effort, and potential adverse effects, and may be justified for a serious outcome but not for a minor one. The individual patient, by contrast, may also care about the relative effect, so presenting both absolute and relative measures supports more informed decisions.

Converting relative risk (RR) to RD requires knowing the baseline event rate in the control group, which is reported in trial results tables. A calculator to facilitate this conversion is available at: <https://evisalud.com/calculadorgrade/>. Once you have this, the NNT can be calculated as follows: $NNT = 1 / RD$. Trials usually report only relative measures; routinely converting them into absolute measures (and assessing whether these effects are meaningful to patients) helps us see beyond the headline figure and supports more informed decision-making.

Rule 7: Assess the risk of bias systematically

Not all RCTs are created equal. The original Cochrane Risk of Bias tool provides a structured framework

for evaluating the internal validity of a trial across five domains: sequence generation (randomization), allocation concealment, blinding of participants and personnel, blinding of outcome assessment, and selective reporting [16]. However, its current revision, RoB 2, reorganizes these into five differently named domains, such as bias arising from the randomization process and bias in selection of the reported result [17].

Here, “systematically” means working through each domain in turn rather than forming a single global impression. For each domain, the question is whether the design or conduct of the trial could have systematically biased the trial results in a particular direction (Table 3).

We recognize that artificial intelligence tools (such as ChatGPT, Gemini, or Claude) are increasingly being used to assess risk of bias. Their value in this setting has not been rigorously established, and they should be used with particular caution by less experienced readers. These tools do not replace a careful, in-depth appraisal of the original study, and their outputs depend on the quality of the input and the user’s understanding. At most, they should be considered supportive aids, never substitutes for critical judgment.

Rule 8: Consider who was (and was not) in the trial

External validity (the applicability of results to populations beyond those studied) is not a function of a trial’s methods but of its participants. Carefully read the inclusion and exclusion criteria. Trials often enroll patients who are younger, healthier, more adherent, and less comorbid than those seen in everyday clinical practice [18].

It helps to separate two questions. The first concerns external validity in the strict sense: are the trial participants (defined by their eligibility criteria and baseline characteristics, such as disease severity and treatment history) similar enough to your patients that

Table 3. Risk-of-bias domains (adapted by the authors from the Cochrane Risk of Bias tool [16])

Bias domain	Key question to ask	Red-flag example
Randomization	Was the allocation truly random?	“Alternate-day allocation” is not randomization.
Blinding of participants and personnel	Were participants and caregivers blinded?	An open-label (i.e., unblinded) design may lead participants to behave differently based on their group assignment (e.g., seeking additional treatments), biasing the measured effect through differential co-interventions between groups.
Blinding of outcome assessment	Were assessors blinded?	The personnel adjudicating a subjective outcome are aware of participants’ allocation to the intervention or control group.
Incomplete data	Is dropout reported and handled?	Substantial or differential loss to follow-up handled without appropriate methods (no universal threshold applies; ≈20% is only a thumb rule).
Reporting	Are all pre-specified outcomes reported?	Primary outcome changed after the trial began without a transparent justification (not all pre-specified changes introduce bias if disclosed and justified).

a comparable effect can be expected? A difference in country or setting alone does not, by itself, imply a different treatment effect; what bears on the expected effect are relevant differences in patients. The second question concerns applicability in practice: can the intervention realistically be delivered in your setting given its resources and constraints?

Examine the baseline characteristics table (typically the first table of any trial report) with the first question in mind, and consider your local resources and feasibility with the second. Together, these comparisons are central to translating evidence into practice.

Rule 9: Think about how limitations affect the credibility of the results

Most trial reports include a limitations section, typically found near the end of the Discussion section. This is often the most intellectually honest passage in the article, and one of the most underread. Authors are expected to declare the constraints of their work; a limitations section that is brief, vague, or absent is itself a red flag ^[19].

Following Rule 7 (Assess the risk of bias systematically), readers will have compiled a list of potential limitations, which can be complemented by those discussed in the paper's Discussion section. For each identified limitation, ask whether it affects the credibility of the results (that is, would a better-designed trial likely show a larger or smaller effect?)

For example, if outcome assessors are unblinded and the primary outcome is subjective (e.g., pain), the apparent benefit of the intervention may be inflated because knowledge of group allocation can influence how assessors elicit, interpret, or record patients' reports of pain, warranting cautious interpretation.

Conversely, some limitations may actually strengthen confidence in positive findings. For instance, in a trial of a new antihypertensive drug, poor adherence in the intervention group would likely dilute the true treatment effect. If the study nevertheless reports a reduction in blood pressure, this suggests that the observed benefit may be robust (and potentially even underestimated).

Rule 10: Place the trial in context before changing practice

No single trial should change clinical practice in isolation. The history of medicine contains multiple examples of interventions that appeared beneficial in a landmark trial and harmful (or with no effect) in subsequent replication: hormone replacement therapy for cardiovascular

protection, tight glycemic control in critical illness, and early revascularization strategies in stable angina, among others ^[20].

Before applying a new finding, ask: Is this trial consistent with the broader body of evidence? Has it been replicated? Are the findings plausible given what is known about the mechanism of action? To answer these questions, consult systematic reviews and clinical practice guidelines, which synthesize and contextualize the available evidence. This step belongs to the "application" aim stated in the Introduction: understanding the underlying evidence allows you to apply guideline recommendations with judgment rather than with uncritical acceptance ^[21].

LIMITATIONS

These rules simplify a complex intellectual process, as high-quality critical appraisal requires a deeper understanding of research methods and statistical principles. As previewed in the Introduction, they also do not cover randomized controlled trial designs beyond the standard two-arm parallel format (such as factorial, crossover, cluster-randomized, and adaptive trials), each of which involves distinct methodological considerations. Moreover, although the CONSORT Statement and the Cochrane Risk of Bias Tool are widely used, their application requires judgment, and reasonable appraisers may disagree in their assessments. Finally, the statistical concepts covered in Rules 5 and 6 are introductory; readers who frequently encounter survival analyses, non-inferiority trials, or Bayesian methods will require more advanced resources beyond the scope of this article.

Accordingly, these rules should be viewed as a starting point rather than a substitute for formal training in research methodology.

CONCLUSIONS

Reading an RCT well is a learnable skill. The ten rules presented here (from framing the study question with PICO to placing results within the broader evidence base) offer a structured, accessible entry point for healthcare students and practitioners regardless of prior methodological experience. The goal is clinical judgment informed by evidence: knowing when a trial's findings are solid enough to act on, when they require corroboration, and when they may not apply to a given patient. In a landscape where evidence is abundant and time scarce, these skills are not optional; they are foundational to safe, effective, and patient-centered care.

Author contributions

ATR conceived the article. All authors contributed to drafting and revising the manuscript and approved the final version for publication.

Funding

This study was self-funded by the authors.

Conflicts of interest statement

The authors declare no conflicts of interest regarding the present study.

Generative Artificial Intelligence use statement

The authors acknowledge the use of artificial intelligence tools during the preparation of this manuscript. Specifically, ChatGPT (OpenAI, San Francisco, CA, USA) and Claude (Anthropic, San Francisco, CA, USA) were

used to assist with language editing, improvement of clarity, and translation of selected sections of the text.

Data availability statement

Not applicable.

Acknowledgments

None.

Supplementary material

None.

ORCID

Alvaro Taype-Rondan: <https://orcid.org/0000-0001-8758-0463>

Sergio Goicochea-Lugo: <https://orcid.org/0000-0002-0487-5547>

REFERENCES

- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-72. doi: [10.1136/bmj.312.7023.71](https://doi.org/10.1136/bmj.312.7023.71).
- Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet*. 2002;359(9305):515-519. doi: [10.1016/S0140-6736\(02\)07683-3](https://doi.org/10.1016/S0140-6736(02)07683-3).
- Sibbald B, Roland M. Understanding controlled trials: why are randomised controlled trials important? *BMJ*. 1998;316(7126):201. doi: [10.1136/bmj.316.7126.201](https://doi.org/10.1136/bmj.316.7126.201).
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926. doi: [10.1136/bmj.39489.470347.AD](https://doi.org/10.1136/bmj.39489.470347.AD).
- Greenhalgh T. How to read a paper: the basics of evidence-based medicine and healthcare. 6th ed. Hoboken: Wiley-Blackwell; 2019.
- Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. doi: [10.1136/bmj.c332](https://doi.org/10.1136/bmj.c332).
- Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. eds. *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. Version 6.4. Cochrane; 2023 [cited 2026 Feb 2]. Available from: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current>
- Critical Appraisal Skills Programme. CASP Randomised Controlled Trial Standard Checklist [Internet]. Oxford: CASP; 2020 [cited 2026 Feb 2]. Available from: <https://casp-uk.net/>
- Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12-A13.
- Chan AW, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586. doi: [10.1136/bmj.e7586](https://doi.org/10.1136/bmj.e7586).
- Dwan K, Gamble C, Williamson PR, Kirkham JJ; Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008;3(8):e3081. doi: [10.1371/journal.pone.0003081](https://doi.org/10.1371/journal.pone.0003081).
- Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet*. 2002;359(9306):614-618. doi: [10.1016/S0140-6736\(02\)07750-4](https://doi.org/10.1016/S0140-6736(02)07750-4).
- Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials*. 2000;21(3):167-189. doi: [10.1016/S0197-2456\(00\)00046-5](https://doi.org/10.1016/S0197-2456(00)00046-5). Erratum in: *Control Clin Trials* 2000 Oct;21(5):526.
- Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat*. 2016;70(2):129-133. doi: [10.1080/00031305.2016.1154108](https://doi.org/10.1080/00031305.2016.1154108).
- Citrome L, Ketter TA. When does a difference make a difference? Interpretation of number needed to treat, number needed to harm, and likelihood to be helped or harmed. *Int J Clin Pract*. 2013;67(5):407-411. doi: [10.1111/ijcp.12142](https://doi.org/10.1111/ijcp.12142).
- Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. doi: [10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928).
- Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898. doi: [10.1136/bmj.l4898](https://doi.org/10.1136/bmj.l4898).
- Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?' *Lancet*. 2005;365(9453):82-93. doi: [10.1016/S0140-6736\(04\)17670-8](https://doi.org/10.1016/S0140-6736(04)17670-8).
- Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. doi: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124).
- Prasad V, Vandross A, Toomey C, Cheung M, Rho J, Quinn S, et al. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clin Proc*. 2013;88(8):790-798. doi: [10.1016/j.mayocp.2013.05.012](https://doi.org/10.1016/j.mayocp.2013.05.012).
- Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet*. 2017;390(10092):415-423. doi: [10.1016/S0140-6736\(16\)31592-6](https://doi.org/10.1016/S0140-6736(16)31592-6).